



Univerzita Palackého
v Olomouci

Trable s p-values

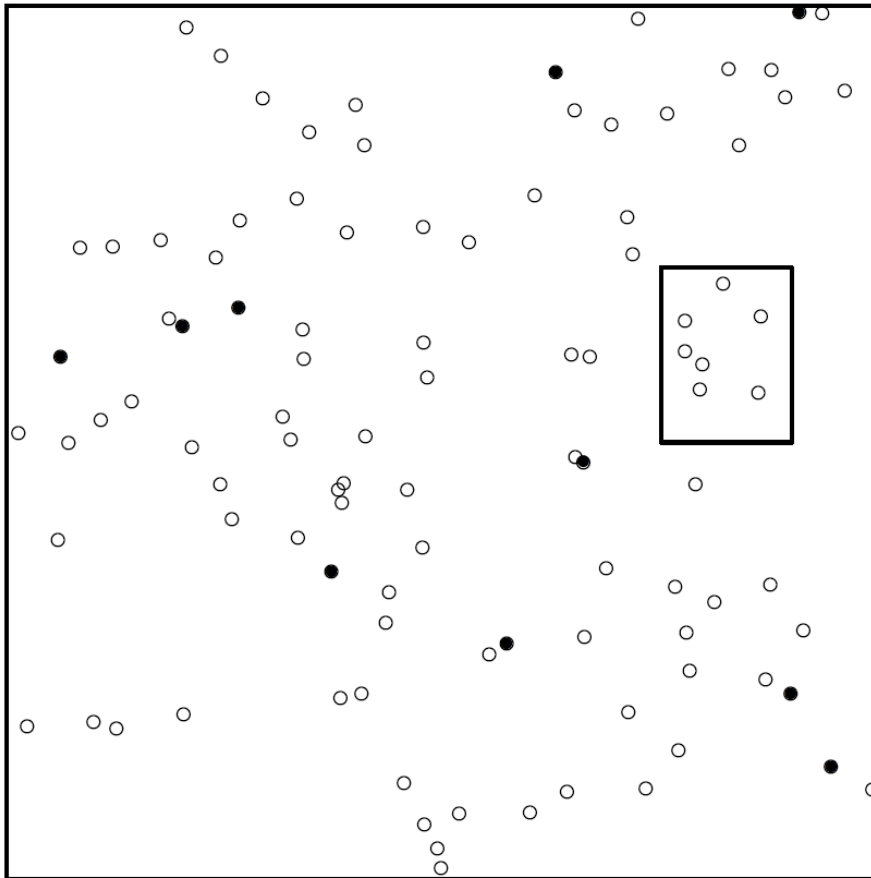
Ondřej Vencálek
PřF UP Olomouc

25. 3. 2025



Univerzita Palackého
v Olomouci

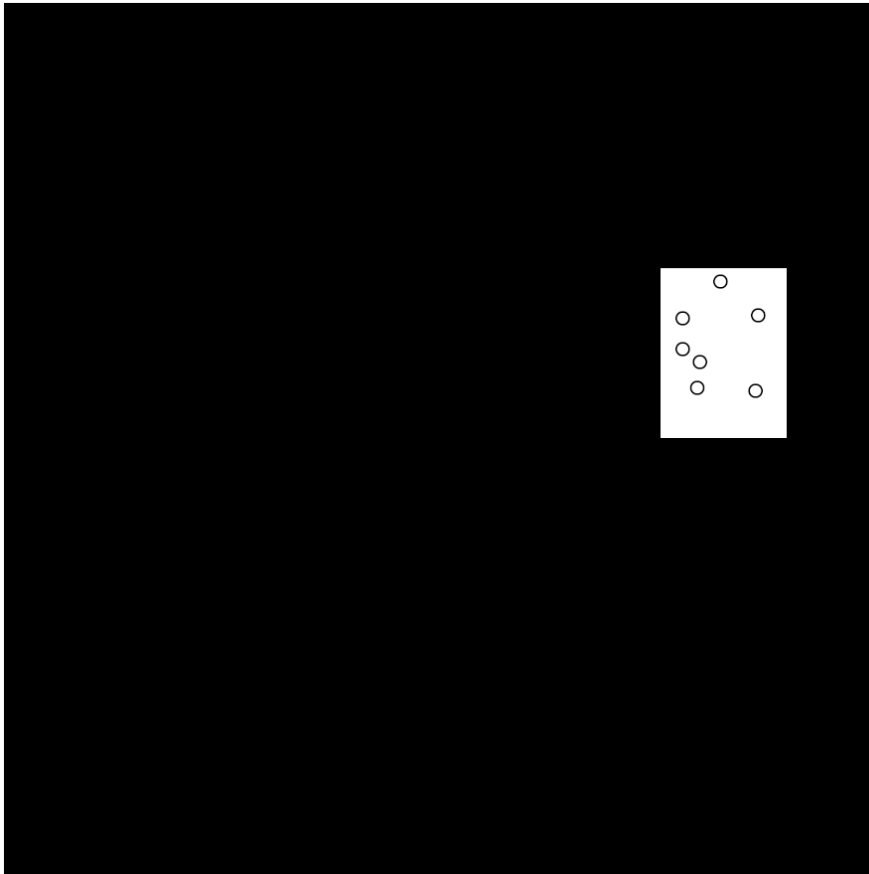
Problém indukce





Univerzita Palackého
v Olomouci

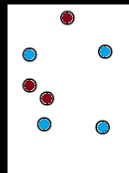
Problém indukce



- Uvědomujme si existenci NEJISTOTY v našich závěrech.
- Umíme se s nejistotou vypořádat, neumíme ji odstranit.



Problém indukce

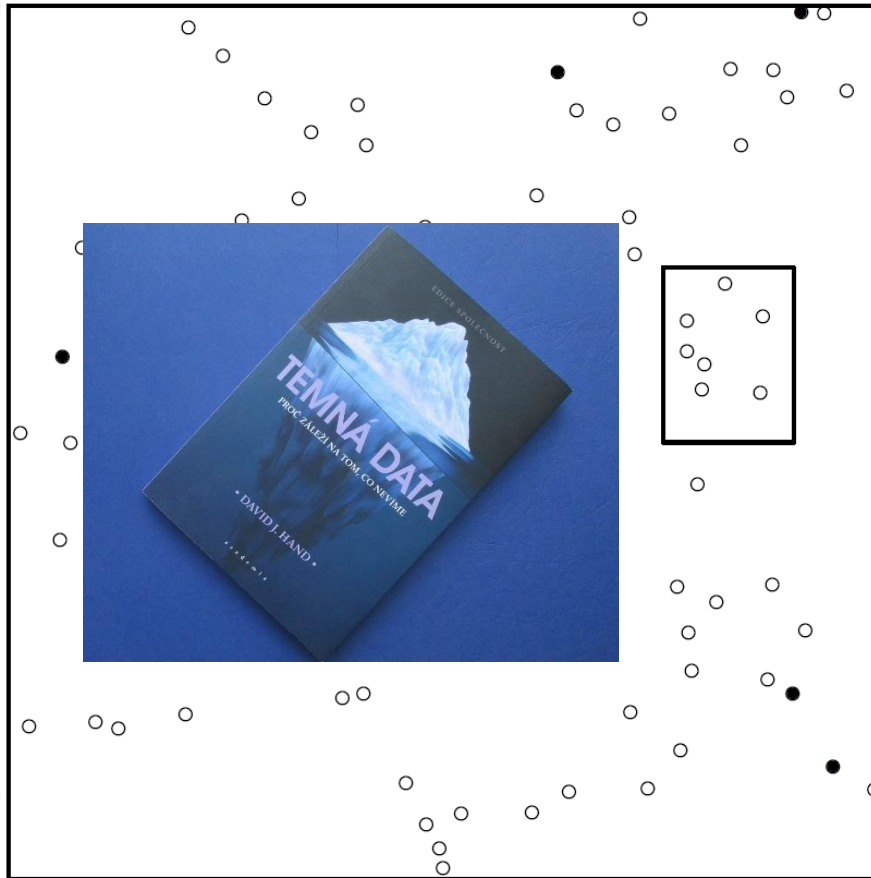


- Světle modré tečky: 4
Tmavě červené tečky: 3
- Otázka:
Můžeme říct, že (v celku) je světle modrých teček více než tmavě červených?
- S jistotou nikoliv
(dokud nepozorujeme celek)
- Jak číst výsledek
„statisticky signifikantní“
a jak „nesignifikantní“.



Univerzita Palackého
v Olomouci

Problém indukce



Hand, D. J. Temná data: proč záleží na tom, co nevíme. Praha: Academia, 2023

Kap.7: Věda a temná data: povaha objevování

- Základním procesem vědy je testování teorií na základě pozorovaných dat, přičemž nesoulad mezi teorií a těmito daty vede k odmítnutí nebo ke změně teorie.
- Kritérium testovatelnosti je to, co odlišuje vědu od pseudovědy.



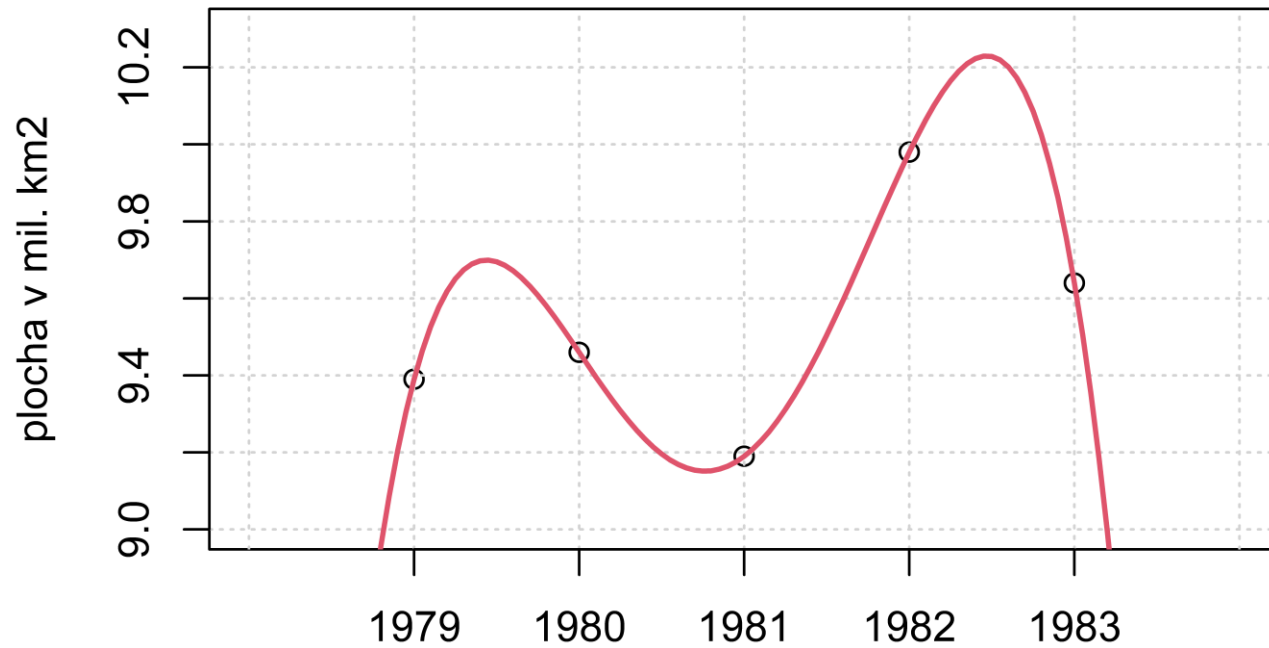
Co dělat a co nedělat

- Explorační analýza (EDA) včetně vizualizace – první krok; podceňovaná; nepřemýšlíme o možnostech zobecnění (zdroj hypotéz)
- Testování hypotéz (spojeno s p-hodnotou) je *nesprávně* používáno jako nástroj *explorační analýzy*.
Problém mnohonásobného testování (p-hacking)
- Trénovací versus testovací datová množina (různé typy cross-validací);
pochopení problému overfitting;
HARKing: Hypothesizing After the Result is Known



Univerzita Palackého
v Olomouci

Overfitting

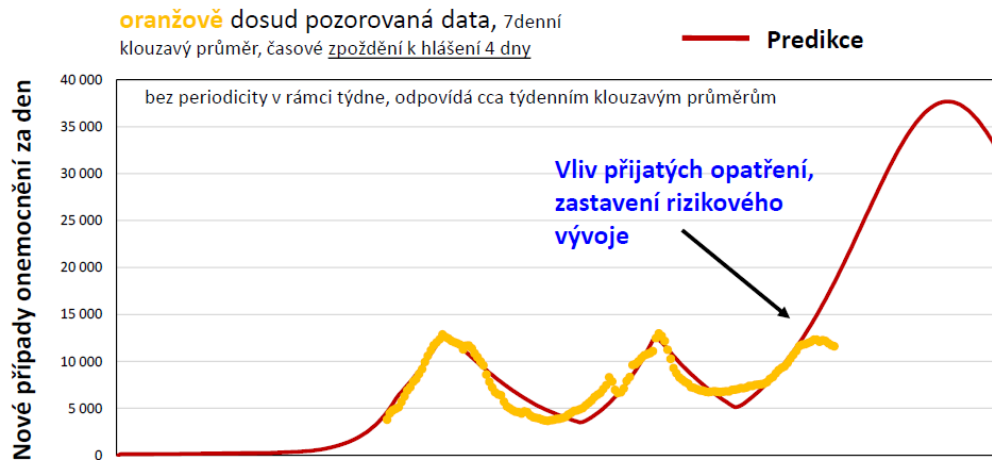


proložení polynomem 4. stupně



Overfilling in real life

SEIR model: scénář pro rozvolnění v březnu a eskalace šíření nákazy Model pokračování výrazného nárůstu pozorovaného v únoru



	červenec	srpen	září	říjen	listopad	prosinec	leden	únor	březen	duben
Nové za měsíc	5 000	8 000	46 000	298 000	199 000	198 000	254 000	292 000	828 000	1 077 000
Kumulativně	17 000	25 000	71 000	369 000	568 000	766 000	1 020 000	1 312 000	2 140 000	3 217 000

Uvolnění od 3.12.

- zvýšení pracovních kontaktů na 80 %
- zvýšení školních kontaktů na 50 %
- zvýšení jiných kontaktů na 100 %
- zvýšení základní reprodukce na původní úroveň

Efekt opatření od 28.12.

- snížení pracovních kontaktů na 40 %
- školní kontakty na 0 %
- snížení jiných kontaktů na 40 %
- snížení základní reprodukce o 10%
- školní kontakty na 30 % od 4.1.

Od 25.1.

- zvýšení základní reprodukce o 70 %
(simulace šíření nakažlivějších forem viru, simulace nedodržování opatření nebo rozvolnění)
(uvedeny podíly normálních kontaktů)

Snímek prezentuje výsledky simulace prostřednictvím epidemiologického modelu SEIR, který zahrnuje vybrané předpoklady a slouží ke zkoumání dopadu změny různých parametrů epidemie. Vzhledem k neurčitostem ve struktuře modelu (například limitované znalosti o skutečné vnímavosti populace k viru a k jeho novým variantám) je nezbytné výsledky brát jako orientační, umožňující zejména celkové srovnávání jednotlivých scénářů, nikoli konkrétní předpověď pro určité období.



p-hacking

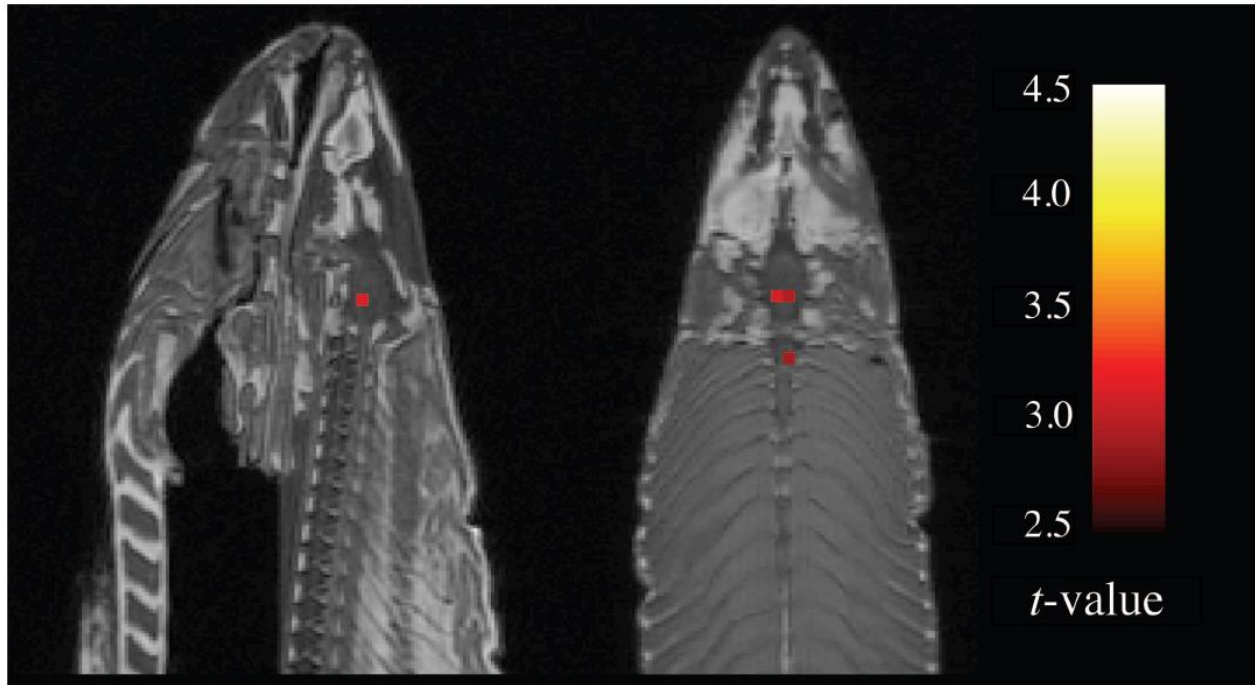


Fig. 1. Sagittal and axial images of significant brain voxels in the task > rest contrast. The parameters for this comparison were $t(131) > 3.15$, $p(\text{uncorrected}) < 0.001$, 3 voxel extent threshold. Two clusters were observed in the salmon central nervous system. One cluster was observed in the medial brain cavity and another was observed in the upper spinal column.



Univerzita Palackého
v Olomouci

TCH
(Toxická chemikálie)



NEMOC

	TCH	nonTCH	
Onemocněl	27	14	41
Neonemocněl	48	67	115
	75	81	156

$$RR = (27 / 75) / (14 / 81) = 2,1$$

95 % CI: (1,2; 3,7)



Univerzita Palackého
v Olomouci

TCH
(Toxická chemikálie)



NEMOC

KOUŘENÍ



Nekuřáci	TCH	nonTCH	
Onemocněl	1	2	3
Neonemocněl	24	48	72
	25	50	75

$$RR = (1 / 25) / (2 / 48) = 1,0$$

95 % CI: (0,1; 10,5)

Kuřáci	TCH	nonTCH	
Onemocněl	26	12	38
Neonemocněl	24	19	43
	50	31	81

$$RR = (26 / 50) / (12 / 31) = 1,3$$

95 % CI: (0,8; 2,3)



Univerzita Palackého
v Olomouci

Něco ke čtení

Hand, D. J. *Temná data: proč záleží na tom, co nevíme*. Praha: Academia, 2023

https://www.statspol.cz/knihovnicka/knihovnicka-Temna_Data/

Taleb, N. N. *Zrádná nahodilost: o skryté roli náhody na trzích a v životě*. Praha: Paseka, 2013. ISBN 978-80-7432-292-1.